“In silico genome analysis and finding a target protein for Mycobacterium Tuberculosis (H37Rv)”

A Project Thesis Submitted in Partial Fulfillment of The

Requirement for the Degree in

Bachelor of Technology

In Biomedical

Submitted by:-

Shraddhananda Biswal

110BM0023

Under the Supervision of:-

Dr. Nandini Sarkar

Assistant Professor

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Odisha-769008

National Institute of Technology, Rourkela

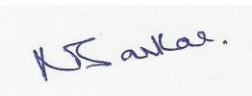
Odisha-769008

CERTIFICATE

This is to certify that the project report entitle “In silico genome analysis and finding a target protein for Mycobacterium Tuberculosis (H37Rv)” submitted by SHRADDHANANDA BISWAL (110BM0023) in the partial fulfillment of the requirement for the degree of the B.Tech in Biomedical Engineering in Department of Biotechnology and Medical Engineering, National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision. To the best of my knowledge the matter embodied in the report has not been submitted to any other Institute/University for any degree.

Date: - 11th May, 2014

Place: - Rourkela



Dr. Nandini Sarkar (Supervisor)

Assistant Professor

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela, Odisha-769008

ACKNOWLEDGEMENTS

I would like to take this opportunity to extent my hearty gratitude to my guide and advisor Dr. Nandini Sarkar, Assistant Professor; Department of Biotechnology and Medical Engineering; National Institute of Technology-Rourkela,Odisha-769008, whose constant guidance and encouragement made the completion of my B.Tech thesis possible.

I would also like to thank Prof. S. Paul; Assistant Professor; Department of Biotechnology and Medical Engineering; National Institute of Technology-Rourkela, Odisha-769008, guiding me maintain the project record book.

I also thank National Institute of Technology Rourkela, for permitting me to utilize the facilities in its laboratories to carry out my experiment.

I would also like to thank Arun Maurya, M.Tech for giving me constant support with all the necessary things involve in this whole project.

Submitted by

Shraddhananda Biswal

Roll No.-110BM0023

Department of Biotechnology & Medical Engineering

National Institute of Technology-Rourkela, Odisha-769008

Table of Contents

List of Figures

List of Tables

Abstract:

Mycobacterium tuberculosis (Mtb) is the bacteria for causing TB in human. H37Rv strain is responsible for this disease. Among the bacterial infectious diseases it is the leading cause of deaths in the world. According to statics it kills 30000 people per year in India. Multi-drug-resistant tuberculosis (MDR TB) shows resistance against these effective drugs- isoniazid and rifampicin. In this approach we have found certain target protein which can give rise to novel drug for this disease. All the metabolic pathways which are not present in the human but present in pathogen are taken as unique pathways. Here there are five such pathways which unique and are present only in bacteria. The genes responsible for these unique pathways were listed and were analyzed for their essentiality for the survival of the pathogen from the DEG (Database of Essential Genes). The essential genes were then blasted against the human genome through BLASTP. The structures and functions of non-homologous proteins were then analyzed.

1. INTRODUCTION:

Tuberculosis is a highly contagious air borne disease. It is believed that that originate from soil. It is postulated that Mtb has originated from mycobacterium bovis, which infects primates and ruminants. During 5000 B.C when man domesticated cattle this strain was introduced into man. This strain underwent a complex process of evolution and host adaptions, inside the human organs to become tubercle bacillus, causing TB [1].

It is one of the leading bacterial diseases. It kills about 1.1 million peoples over the world, an additional 0.35 million deaths were due to HIV-associated tuberculosis. India shares a big portion of this figure [2]. The existing drugs, have several shortcomings, the most important is the emergence of drug resistance strain developed due to this drugs and as a result frontline drugs are inactive. Secondly, they make the patient incompliance. Another important problem with most of the existing antimycobacterials is their inability to act upon latent forms of the bacillus. In addition to these problems, the vicious interactions between the HIV (human immunodeficiency virus) and TB have led to further challenges for antitubercular drug discovery [3]. The mycobacterial cell wall envelope is thick, rigid, and waxy and consists of inner lipid bilayer plasma membrane. The cell wall is formed by peptidoglycan-arabinogalactan polymers in periplasmic space with outer lipid enriched in mycolic acids covalently Linked To The Arabinogalactan Layer [4].

1.1. Mode of Infection and Symptoms:

Mycobacterium tuberculosis can only infect a healthy person from an infected person. The mode of transfer of this bacteria are coughing, sneezing and talking of infected person can release the bacteria into the surrounding air and the people breathing this air get infected. It primarily infect the lungs but

also has severe effects on central nervous system, lymphatic system, circulatory system among other organs [5].

When this disease infect humans it shows a range of symptoms such as severe cough, fever, chills, night sweets, loss of appetite, severe weight loss, blood in sputum,etc. A person who does not show any symptoms of the active disease is referred as inactive TB. Someone with a healthy immune system has 10% lifetime chances for reactivating this inactive bacterium into active symptoms of TB. But if the person is suffering from AIDS or other weakness diseases that suppress the immune system, then the chance of reactivating increases to 10% each year. The other weakness diseases include:

1. Diabetes.
2. Head or neck cancer.
3. Kidney disease.
4. Long term steroid use.
5. Malnutrition
6. Medications that suppress the immune system, such as anticancer medications (e.g., cyclosporine, tacrolimus).

1.2 Earlier Therapeutic Approach:

In early 1930s and 1940s antibiotics like Penicillin and Sulfa drugs were used, but it soon became evident that this bacterium was resistant to these drugs. In the year 1943, Selman Waksman found Streptomycin as anti-tuberculosis agent. After few years until 1948, it was found that a new strain was developed which was resistant to Streptomycin. Then two new drugs para-aminosalicylic acid and thiacetazone were developed. When these two drugs in combination with Streptomycin were used the antibiotics resistance strain was reduced significantly. In 1951 isoniazid (isonicotinic acid hydrazide)

was proved be best clinical outcomes but for few years. Followed by isoniazid, there were few more new drugs emerged for mutant strain such as pyrazinamide (1952), cycloserine (1952), ethionamide (1956), rifampicin (1957) and ethambutol (1962).

So here the first line of drugs are: isonazide amd rimpicin, pyrazinamide, ethambutol and streptomycin. According to DOTS (directly observed therapy, short-course), the sputum is tested and first line of drugs are being administered. The disadvantages of this method are :

i. The treatment regime is 6-9 months.
ii. Side-effects such as vomiting, dizziness and skin rashes to drug induced hepatitis.

The irregular intake of medicine causes MDR-TB (Multiple Drug Resistant TB), and here the first line of drugs fail [6].

1.3 Objective:

The objective of this project is to identify a gene of unique metabolism pathways which can act a target for mycobacterium tuberculosis. That gene should be non-homologous to host(here human).

2. LITERATURE REVIEW:

NCBI:

National Center for Biotechnology Information (NCBI), it is a part of the United States National Library of Medicine. It contains the tools such as BLAST (for sequence similarity), GenBank (provide the nucleotide sequence), PubMed [7].

BLASTP: Basic Local Alignment Search Tool. It is a tool used to compare a query sequence of nucleotides with a database of sequence to identify the similarity between the two sequences above a threshold [8]. BLASTP: Here the amino acid sequence is compared. As a result, it gives clear information about the similarity of genes and which is related to some protein**.**

KEGG:

Kyoto Encyclopedia of Genes and Genomes is a set of database of biological pathways, diseases, drugs, chemical substances []. Used for analysis of genomics, metagenomics and metabolomics. It is an accumulation of pathway maps incorporating numerous substances including genes, proteins, RNAs, substance mixes, glycans, and compound responses, and in addition disease genes and targets, which are put away as individual entrances in alternate databases of KEGG[9].

DEG:

Database of Essential Genes, is a database and provide tool to analysis the essentiality of the gene [10]. Essential genes are those genes of an organism that are thought to be critical for its survival of the organism.

UNIPORT:

It gives information of the gene about the function, sequence and location in the cell. UniProt Knowledgebase is a protein database partially curated by experts, consisting of two sections: UniProtKB/Swiss-Prot (containing reviewed, manually annotated entries) and UniProtKB/ TrEMBL (containing unreviewed, automatically annotated entries) [11].

3. MATERIALS METHODS:

3.1 Selection of Unique Pathways:

KEGG data base is used to select the unique pathways from human and mtb. It has been found that there are some unique pathways found in the bacteria [5].

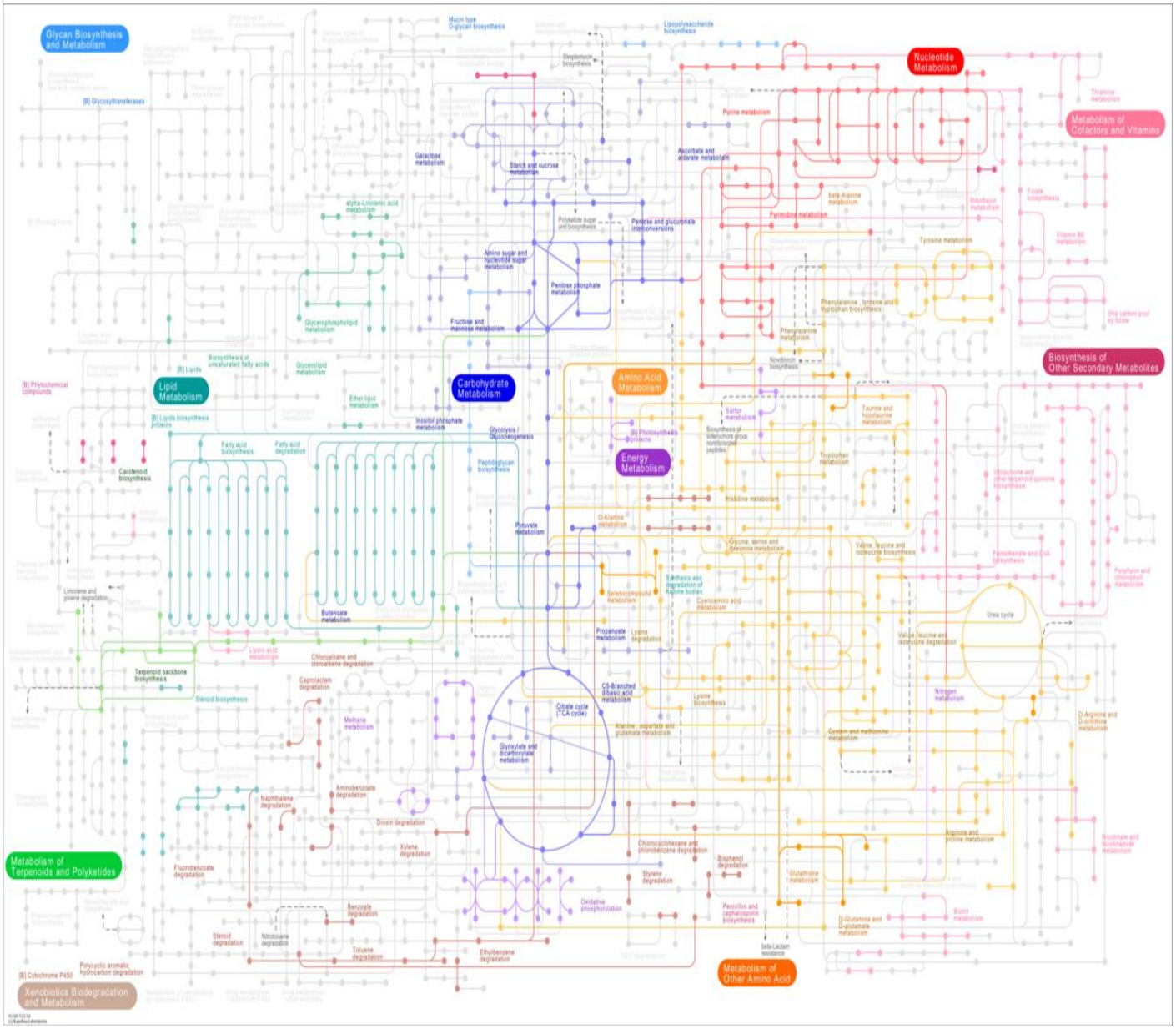


Fig: 1 (Pathways of Mtb)

The unique pathways are listed here:

1. Carbohydrate metabolism
   (a) c5-Branched dibasic acid metabolism.
2. Energy metabolism
   (a) Carbon fixation pathway in prokaryotes.
   (b) Methane metabolism.
3. Lipopolysaccharide biosynthesis.
4. Peptidoglycan biosynthesis.

Here we have selected four (excluding carbon fixation in prokaryotes) among the five pathways, because here we have concentrated towards the genes of plasma membrane/ cell wall for vaccination purpose.

3.2 List of genes:

Steps to be followed to get the list of all genes are:

Open KEGG website→metabolism pathway→select the pathway listed above→in the dropdown box select the mycobacterium tuberculosis H37Rv→click on Go→ pathway entry.

A page will open, which will contain the list of genes and shown as:

Gene	
	Rv1820 ilvG; acetolactate synthase large subunit IlvG [KO:K01652] [EC:2.2.1.6]
	Rv3003c ilvB1; acetolactate synthase large subunit IlvB [KO:K01652] [EC:2.2.1.6]
	Rv3470c ilvB2; acetolactate synthase large subunit [KO:K01652] [EC:2.2.1.6]
	Rv3509c ilvX; acetohydroxyacid synthase large subunit [KO:K01652] [EC:2.2.1.6]
	Rv3002c ilvN; acetolactate synthase small subunit [KO:K01653] [EC:2.2.1.6]
	Rv0951 sucC; succinyl-CoA ligase subunit beta [KO:K01903] [EC:6.2.1.5]
	Rv0952 sucD; succinyl-CoA ligase subunit alpha [KO:K01902] [EC:6.2.1.5]
	Rv2988c leuC; 3-isopropylmalate dehydratase large subunit [KO:K01703] [EC:4.2.1.35 4.2.1.33]
	Rv2987c leuD; 3-isopropylmalate dehydratase small subunit [KO:K01704] [EC:4.2.1.35 4.2.1.33]

Fig: 2 snapshot of KEGG database contains list of genes.

All the genes involved in these pathways are listed down:

1. C5-Branched dibasic acid metabolism

1.	mtu:Rv0951	sucC; Probable succinyl-CoA synthetase (beta chain) SucC (SCS-beta); K01903 succinyl-CoA synthetase
2.	mtu:Rv0952	sucD; Probable succinyl-CoA synthetase (alpha chain) SucD (SCS-alpha); K01902 succinyl-CoA synthetas
3.	mtu:Rv1820	ilvG; Probable acetolactate synthase IlvG (acetohydroxy-acid synthase)(ALS); K01652 acetolactate syn
4.	mtu:Rv2987c	leuD; Probable 3-isopropylmalate dehydratase (small subunit) LeuD

		(isopropylmalate isomerase) (alpha
5.	mtu:Rv2988c	leuC; Probable 3-isopropylmalate dehydratase (large subunit) LeuC (isopropylmalate isomerase)
6.	mtu:Rv3002c	ilvN; Probable acetolactate synthase (small subunit) IlvN (acetohydroxy-acid synthase) (AHAS) (ALS);
7.	mtu:Rv3003c	ilvB1; Acetolactate synthase (large subunit) IlvB1 (acetohydroxy-acid synthase); K01652 acetolactate
8.	mtu:Rv3470c	ilvB2; Probable acetolactate synthase (large subunit) IlvB2 (AHAS) (acetohydroxy-acid synthase large

2. Methane metabolism:

1. Rv0761c	adhB; Possible zinc-containing alcohol dehydrogenase NAD dependent AdhB [KO:K00121] [EC:1.1.1.1 1.1.1.284]
2. Rv3086	adhD; Probable zinc-type alcohol dehydrogenase AdhD (aldehyde reductase) [KO:K00121] [EC:1.1.1.11.1.1.284]
3. Rv0374c	Probable carbon monoxyde dehydrogenase (small chain); K03518 carbon-monoxide dehydrogenase small subunit [EC:1.2.99.2] [KO:K03518] [EC:1.2.99.2]

4. Rv0375c	Probable carbon monoxyde dehydrogenase (medium chain); K03519 carbon-monoxide dehydrogenase medium subunit [EC:1.2.99.2] [KO:K03519] [EC:1.2.99.2]
5. Rv0373c	Probable carbon monoxyde dehydrogenase (large chain); K03520 carbon-monoxide dehydrogenase large subunit [EC:1.2.99.2] [KO:K03520] [EC:1.2.99.2]
6. Rv0070c	glyA2; Serine hydroxymethyltransferase GlyA2 (serine methylase 2) (SHMT 2) [KO:K00600] [EC:2.1.2.1]
7. Rv1093	glyA1; Serine hydroxymethyltransferase 1 GlyA1 [KO:K00600] [EC:2.1.2.1]
8. Rv1023	eno; Probable enolase Eno [KO:K01689] [EC:4.2.1.11]
9. Rv1240	mdh; Probable malate dehydrogenase Mdh [KO:K00024] [EC:1.1.1.37]
10. Rv0363c	fba; Probable fructose-bisphosphate aldolase Fba [KO:K01624] [EC:4.1.2.13]
11. Rv1099c	glpX; Fructose 1,6-bisphosphatase GlpX [KO:K02446] [EC:3.1.3.11]
12. Rv3010c	pfkA; Probable 6-phosphofructokinase PfkA (phosphohexokinase) (phosphofructokinase) [KO:K00850] [EC:2.7.1.11]
13. Rv2029c	pfkB; 6-phosphofructokinase PfkB (phosphohexokinase) (phosphofructokinase) [KO:K16370] [EC:2.7.1.11]
14. Rv0409	ackA; Probable acetate kinase AckA (acetokinase) [KO:K00925] [EC:2.7.2.1]
15. Rv0408	pta; Probable phosphate acetyltransferase Pta (phosphotransacetylase) [KO:K13788] [EC:2.3.1.8]

16. Rv3667	acs; Acetyl-coenzyme A synthetase Acs (acetate--CoA ligase) (acetyl-CoA synthetase) (acetyl-CoA synthase) (acyl-activating enzyme) (acetate thiokinase) (acetyl-activating enzyme) (acetate--coenzyme A ligase) (acetyl-coenzyme A synthase) [KO:K01895] [EC:6.2.1.1]
17. Rv0489	gpm1; Probable phosphoglycerate mutase 1 Gpm1 (phosphoglyceromutase) (PGAM) (BPG-dependent PGAM) [KO:K01834] [EC:5.4.2.11]
18. Rv2419c	gpgP; Glucosyl-3-phosphoglycerate phosphatase GpgP [KO:K15634] [EC:5.4.2.12]
19. Rv3214	gpm2; Possible phosphoglycerate mutase Gpm2 (phosphoglyceromutase) (PGAM) (BPG-dependent PGAM) [KO:K15634] [EC:5.4.2.12]
20. Rv3837c	Probable phosphoglycerate mutase (phosphoglyceromutase) (phosphoglycerate phosphomutase); K15634 probable phosphoglycerate mutase [EC:5.4.2.12] [KO:K15634] [EC:5.4.2.12]
21. Rv2228c	Multifunctional protein Has RNASE H; K15634 probable phosphoglycerate mutase [EC:5.4.2.12] [KO:K15634] [EC:5.4.2.12]
22. Rv0728c	serA2; Possible D-3-phosphoglycerate dehydrogenase SerA2 (phosphoglycerate dehydrogenase) (PGDH) [KO:K00058] [EC:1.1.1.95]
23. Rv2996c	serA1; Probable D-3-phosphoglycerate dehydrogenase SerA1 (PGDH) [KO:K00058] [EC:1.1.1.95]
24. Rv0884c	serC; Possible phosphoserine aminotransferase SerC (PSAT) [KO:K00831] [EC:2.6.1.52]
25. Rv3042c	serB2; Probable phosphoserine phosphatase SerB2 (PSP) (O-

	phosphoserine phosphohydrolase) (pspase) [KO:K01079] [EC:3.1.3.3]
26. Rv2983	Conserved hypothetical alanine rich protein; K14941 2-phospho-L-lactate guanylyltransferase [EC:2.7.7.68] [KO:K14941] [EC:2.7.7.68]
27. Rv3261	fbiA; Probable F420 biosynthesis protein FbiA [KO:K11212] [EC:2.7.8.28]
28. Rv3262	fbiB; Probable F420 biosynthesis protein FbiB [KO:K12234] [EC:6.3.2.34 6.3.2.31]

3.Lipopolysaccharide biosynthesis:

1. Rv2611c	Probable acyltransferase; K02517 lipid A biosynthesis lauroyl acyltransferase [EC:2.3.1.-] [KO:K02517] [EC:2.3.1.-]
2. Rv0113	gmhA; Probable sedoheptulose-7-phosphate isomerase GmhA (phosphoheptose isomerase) [KO:K03271] [EC:5.3.1.28]
3. Rv0114	gmhB; Possible D-alpha,beta-D-heptose-1,7-biphosphate phosphatase GmhB (D-glycero-D-manno-heptose 7-phosphate kinase) [KO:K03273] [EC:3.1.3.83 3.1.3.82]
4. Rv0115	hddA; Possible D-alpha-D-heptose-7-phosphate kinase HddA [KO:K07031] [EC:2.7.1.168]

4 .**Peptidoglycan biosynthesis**:

1. Rv1315	murA; Probable UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA [KO:K00790] [EC:2.5.1.7]
2. Rv0482	murB; Probable UDP-N-acetylenolpyruvoylglucosamine reductase MurB (UDP-N-acetylmuramate dehydrogenase) [KO:K00075] [EC:1.3.1.98]
3. Rv2152c	murC; Probable UDP-N-acetylmuramate-alanine ligase MurC [KO:K01924] [EC:6.3.2.8]
4. Rv2155c	murD; Probable UDP-N-acetylmuramoylalanine-D-glutamate ligase MurD [KO:K01925] [EC:6.3.2.9]
5. Rv2158c	murE; Probable UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase MurE [KO:K01928] [EC:6.3.2.13]
6. Rv2981c	ddlA; Probable D-alanine--D-alanine ligase DdlA (D-alanylalanine synthetase) (D-ala-D-ala ligase) [KO:K01921] [EC:6.3.2.4]
7. Rv2157c	murF; Probable UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase MurF [KO:K01929] [EC:6.3.2.10]
8. Rv2136c	Possible conserved transmembrane protein; K06153 undecaprenyl-diphosphatase [EC:3.6.1.27] [KO:K06153] [EC:3.6.1.27]
9. Rv2156c	murX; Probable phospho-N-acetylmuramoyl-pentappeptidetransferase MurX [KO:K01000] [EC:2.7.8.13]
10. Rv2153c	murG; Probable UPD-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol-N-acetylglucosamine transferase MurG [KO:K02563] [EC:2.4.1.227]

11. Rv3910	Probable conserved transmembrane protein; K03980 virulence factor [KO:K03980]
12. Rv0016c	pbpA; Probable penicillin-binding protein PbpA [KO:K05364] [EC:2.4.1.129]
13. Rv2163c	pbpB; Probable penicillin-binding membrane protein PbpB [KO:K03587]
14. Rv2911	dacB2; Probable penicillin-binding protein DacB2 (D-alanyl-D-alanine carboxypeptidase) (DD-peptidase) (DD-carboxypeptidase) (PBP) (DD-transpeptidase) (serine-type D-ala-D-ala carboxypeptidase) (D-amino acid hydrolase) [KO:K07258] [EC:3.4.16.4]
15. Rv3330	dacB1; Probable penicillin-binding protein DacB1 (D-alanyl-D-alanine carboxypeptidase) (DD-peptidase) (DD-carboxypeptidase) (PBP) (DD-transpeptidase) (serine-type D-ala-D-ala carboxypeptidase) (D-amino acid hydrolase) [KO:K07258] [EC:3.4.16.4]
16. Rv3627c	hypothetical protein; K07259 D-alanyl-D-alanine carboxypeptidase / D-alanyl-D-alanine-endopeptidase (penicillin-binding protein 4) [EC:3.4.16.4 3.4.21.-] [KO:K07259] [EC:3.4.21.- 3.4.16.4]

3.3 Selection of essential genes:

The essential genes are common to cells and are considered as foundation of life. The amino acid sequence are compared by using BLASTP of DEG. Importance of the above are analyzed through **DEG (Database of Essential Genes)**, cut off score was set greater than 100, due to the specificity of the enzyme for the existence of the bacterial.

3.4 USE OF BLAST:

Here BLASTP is being used discard the homologous genes. Homologous genes are genes that are present in the human being and plays important role. If we target the homologous genes than this will produce adverse effect in the biological process of the human being. So the identity below 15 % can be taken as target and expected value 0.005. As we are targeting the multi resistant strain it would be better to target the most conserved domain of the bacteria.

4. RESULT AND DISCUSSION:

4.1 List of Pathways with Figure:

The unique pathway of KEGG website are as follows:

1. c5-Branched dibasic acid metabolism.

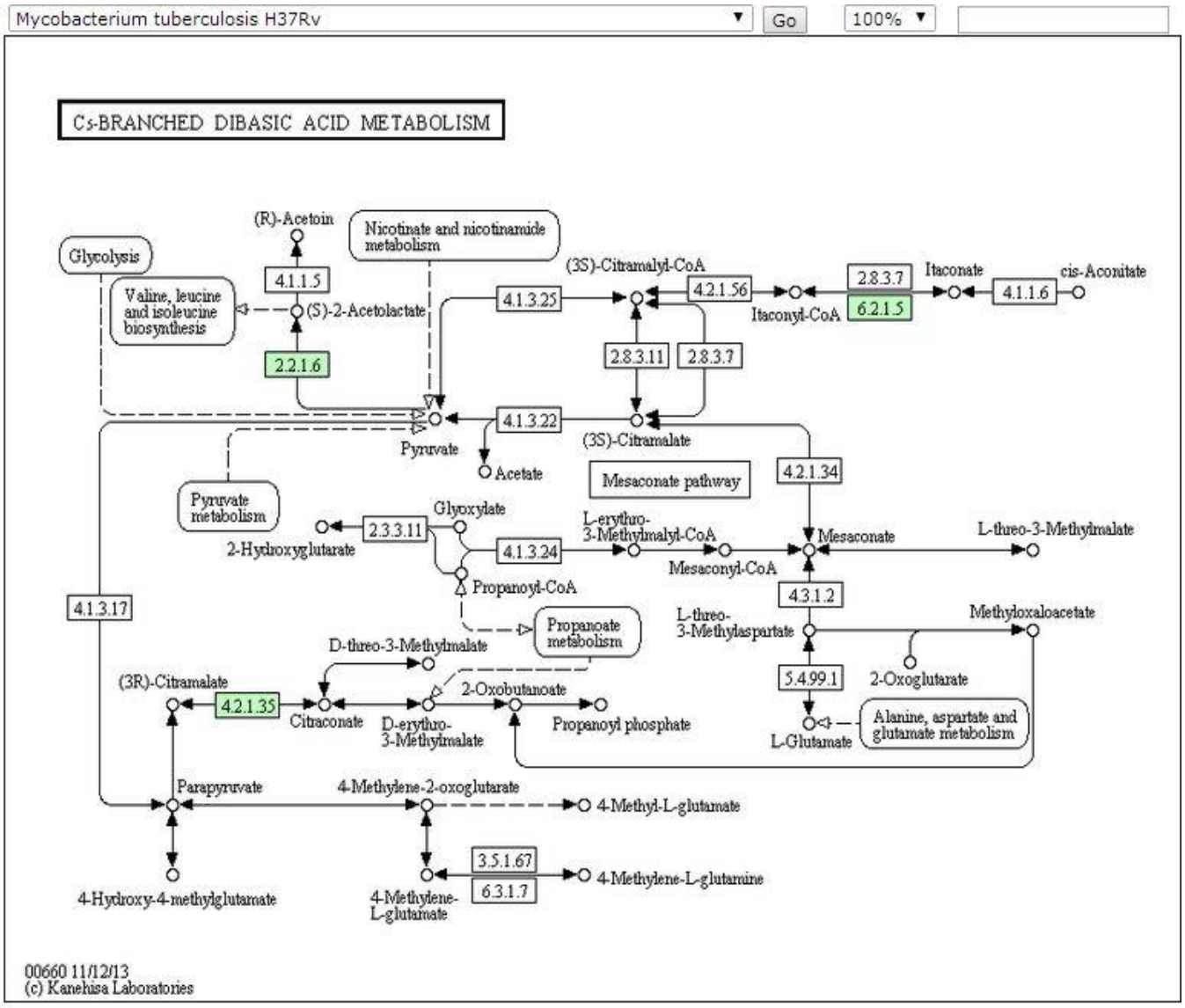


Fig:3

Methane metabolism:

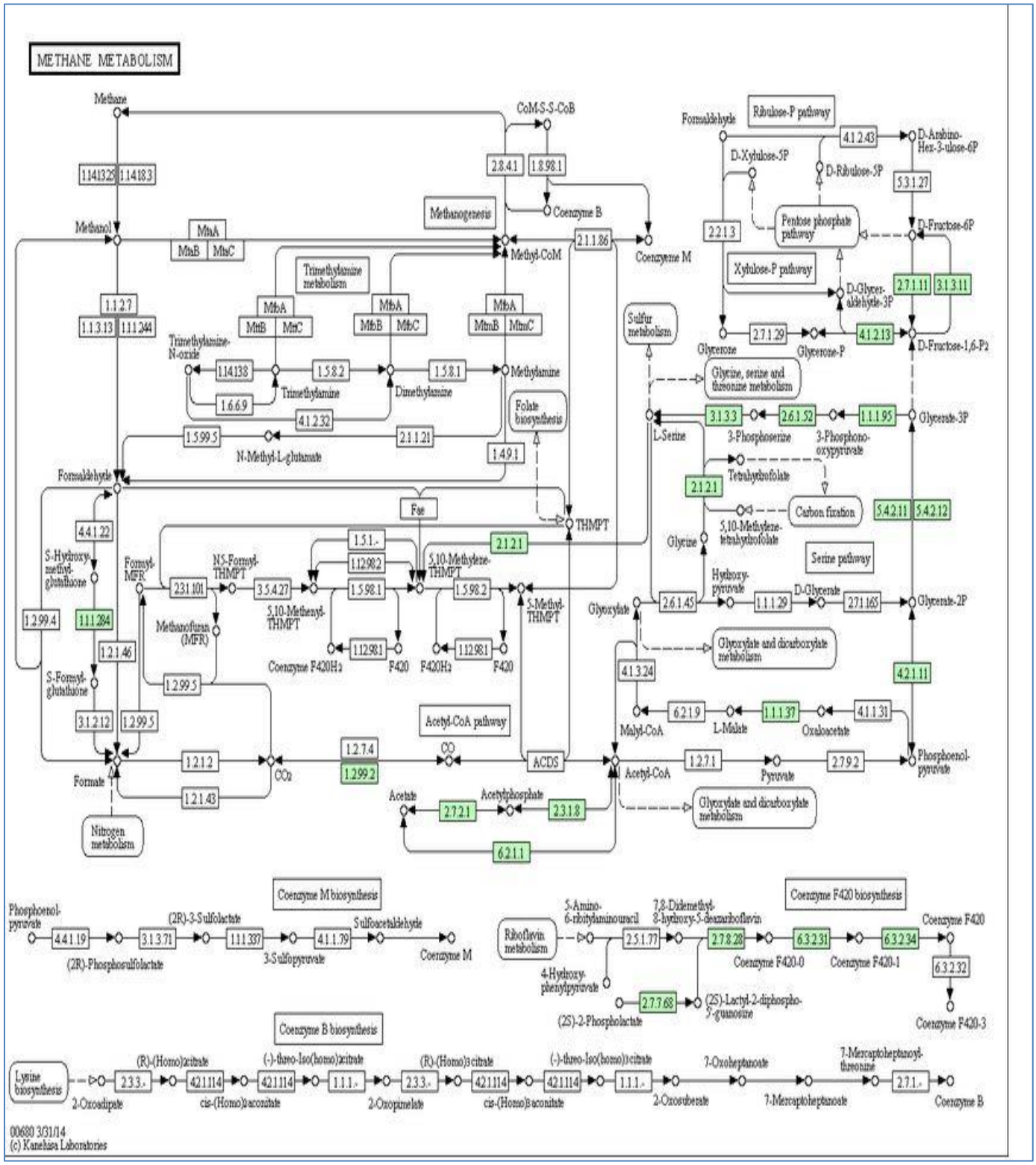


Fig:4

2. Lipopolysaccharide biosynthesis:

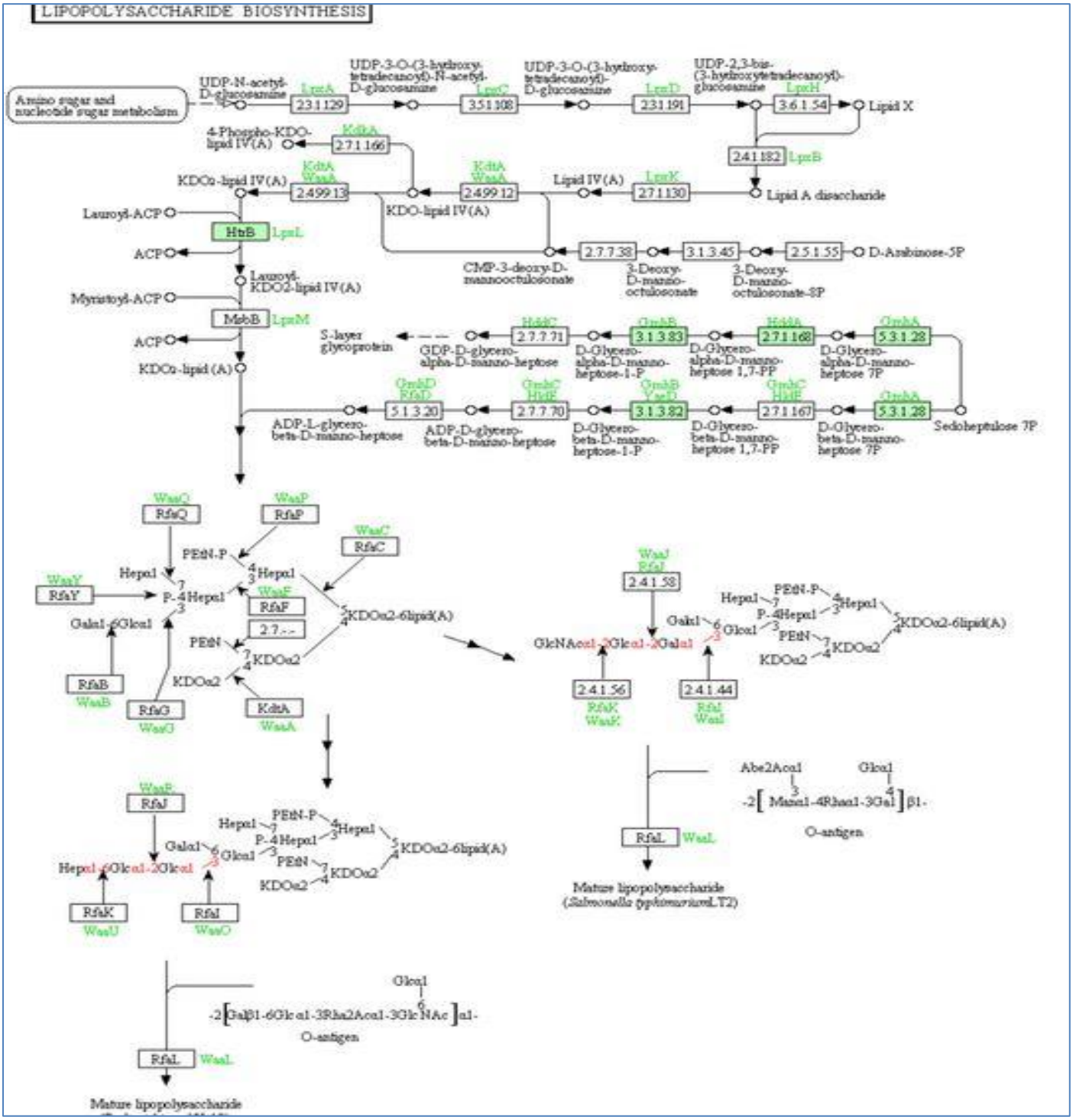


Fig: 5

3. Peptidoglycan biosynthesis:

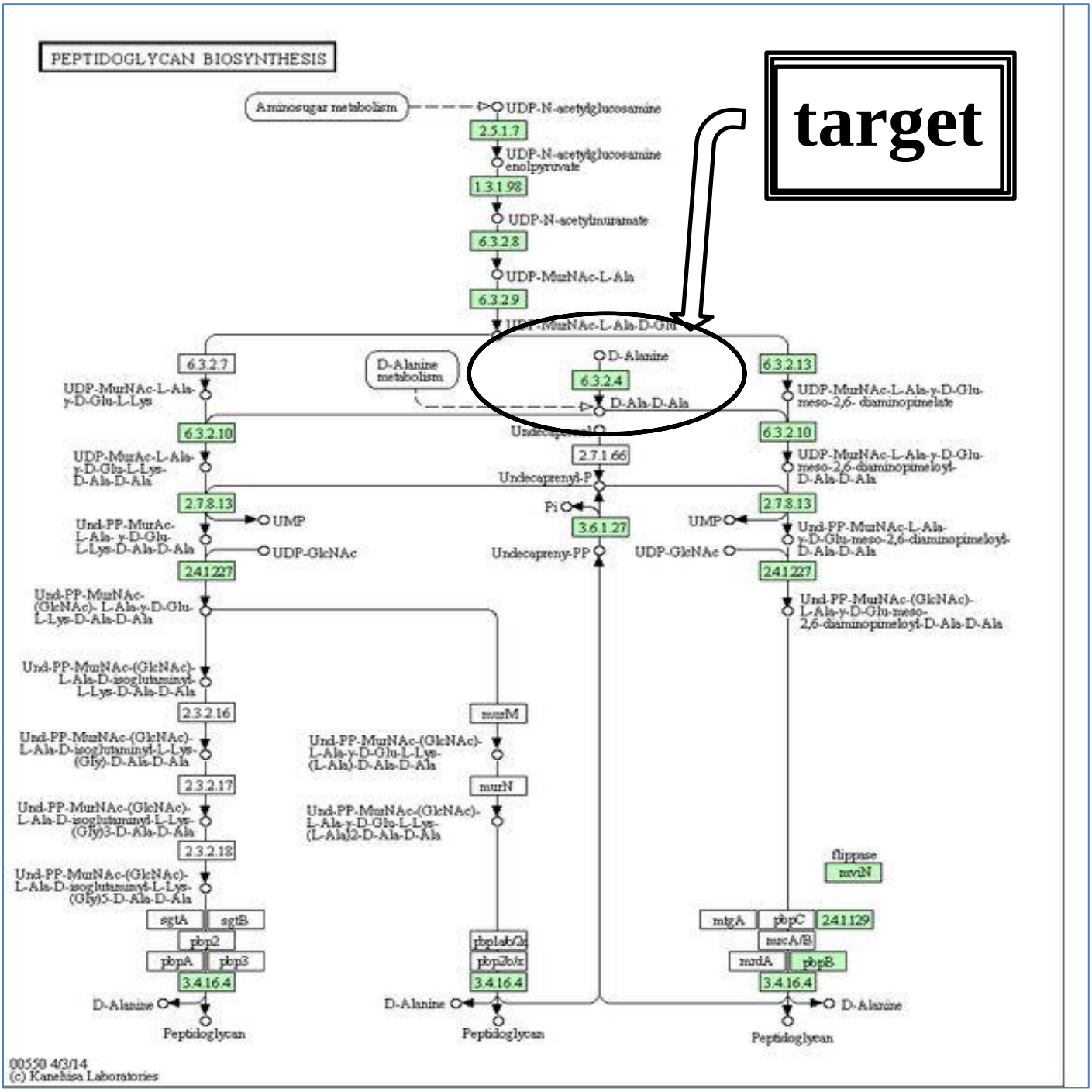


Fig:6

4.2 SELECTION OF ESSENTISL GENES:

The essential genes are short out though the DEG database. The parameters that is being used to short out the genes is shown below, score greater than 100(i.e. 500), expected value is $1 \times e^{0.5}$. This is taken due to specificity of the gene toward the metabolism process.

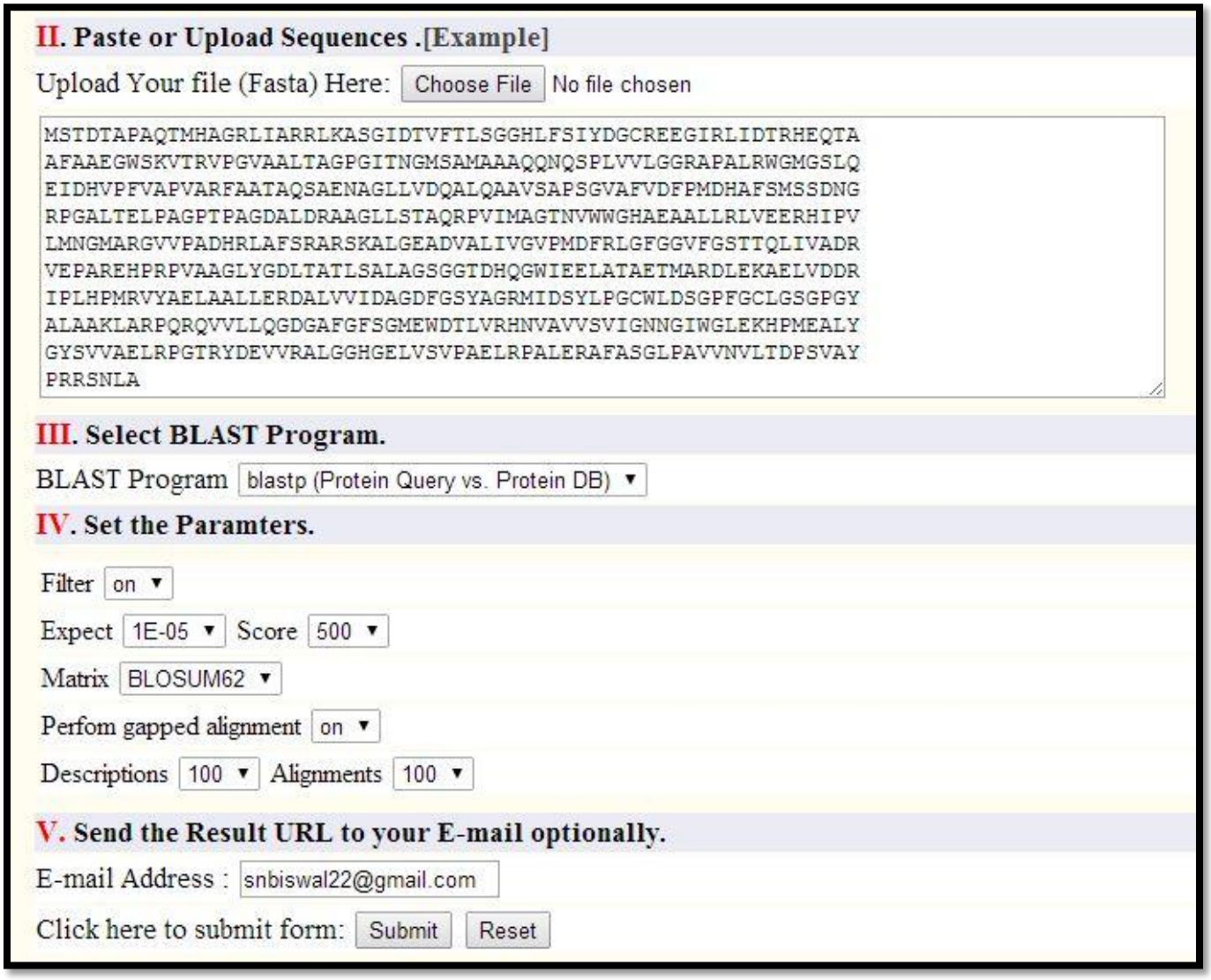


Fig: 7 Parameters for the DEG

List of Essential Genes are:

After submitting all the amino acid sequence of the genes of selected four pathways are retrieved from the DEG database or through mail. All the essential genes are shown below.

Gene Entry	Gene Name
1. **Rv1820**	ilvG
2. **Rv0951**	sucC
3. **Rv0761c**	adhB
4. **Rv0373c**	
5. **Rv0070c**	glyA2
6. **Rv1023**	Eno
7. **Rv1240**	Mdh
8. **Rv0363c**	Fba
9. **Rv2029c**	pfkB
10. **Rv0409**	ackA
11. **Rv0408**	Pta
12. **Rv3667**	Acs
13. **Rv0489**	gpm1

14. Rv0728c	serA2
15. Rv2611c	
16. Rv0113	gmhA
17. Rv0114	gmhB
18. Rv0482	murB
19. Rv2152c	murC
20. Rv2155c	murD
21. Rv2158c	murE
22. Rv2981c	dd1A
23. rv2157c	murF
24. Rv2136c	
25. Rv2156c	murX
26. Rv2153c	
27. Rv3910	
28. Rv0016c	PbpA
29. Rv2163c	PbpA
30. Rv1315	murA

4.3 BLAST:

Result of the BLASTP of human genome against the shortlisted 30 genes. The threshold was set to 0.005 and identity should be less than 35%.

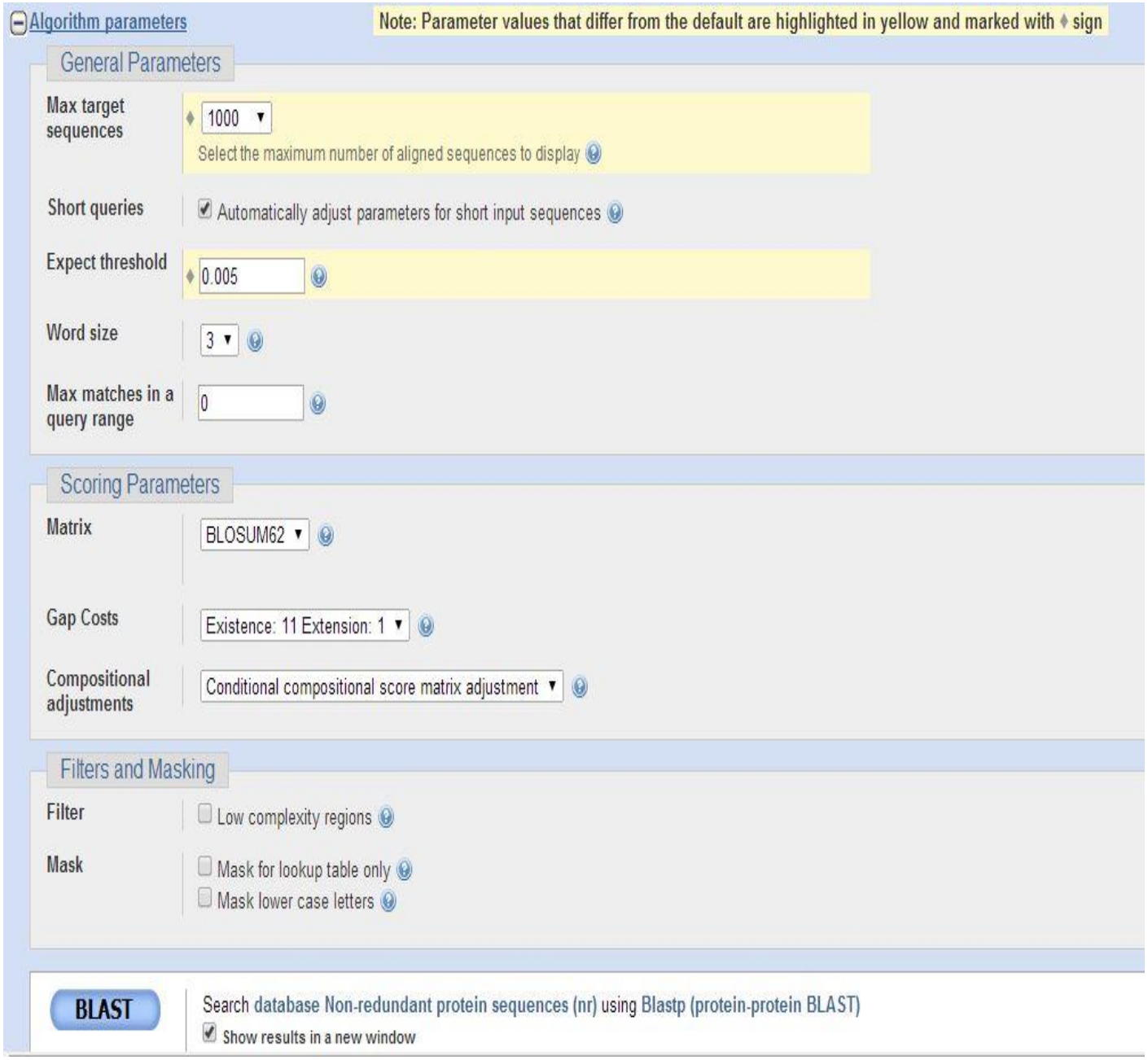
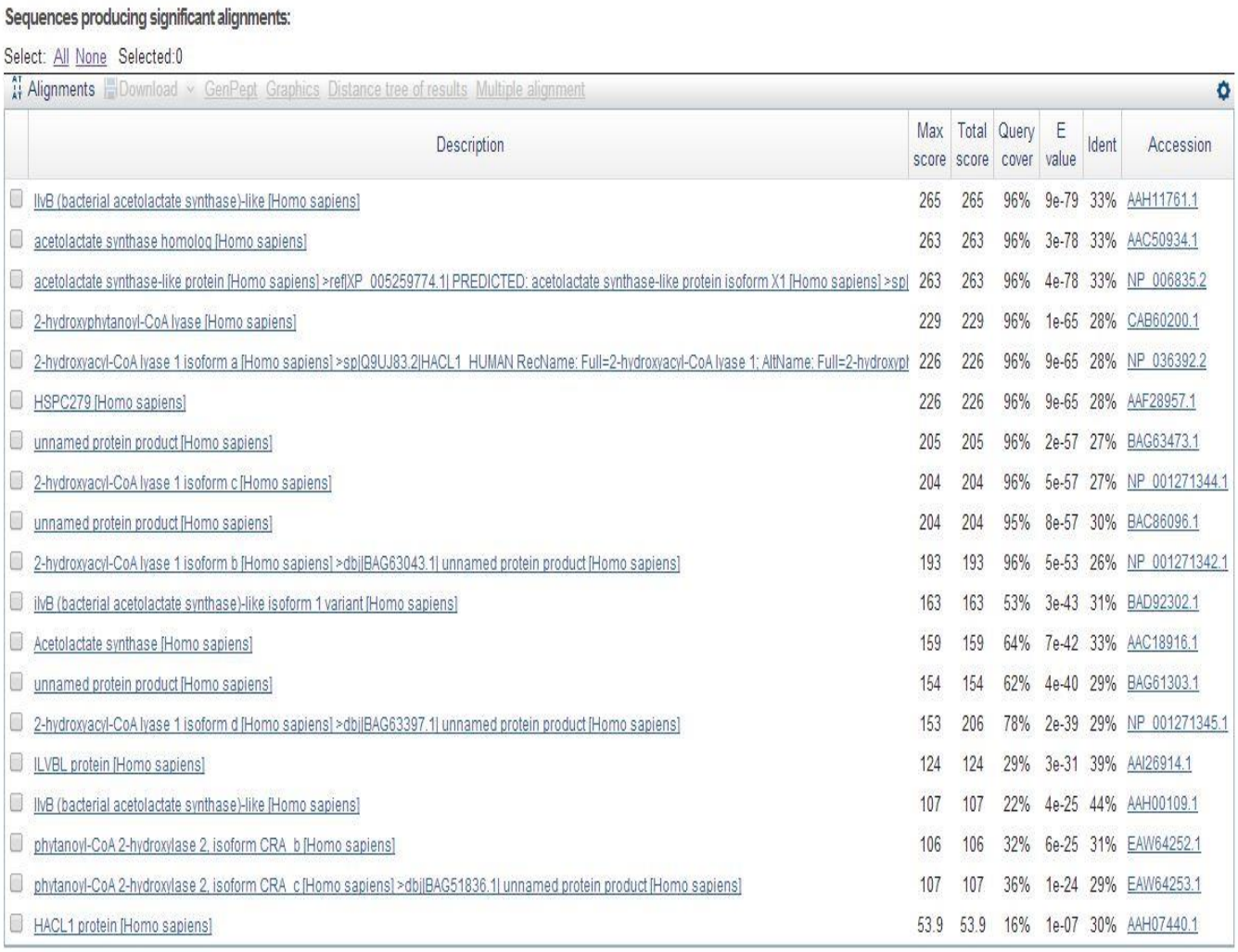


Fig: 8 Parameters for BLASTP

4.3.1 Result of the BLASTP:

After pasting the amino acid sequence in the box provided, setting the parameters as shown above, then click on BLAST. A page will open after few second as such:



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
IlvB (bacterial acetolactate synthase)-like [Homo sapiens]	265	265	96%	9e-79	33%	AAH11761.1
acetolactate synthase homolog [Homo sapiens]	263	263	96%	3e-78	33%	AAC50934.1
acetolactate synthase-like protein [Homo sapiens] >ref|XP_005259774.1| PREDICTED: acetolactate synthase-like protein isoform X1 [Homo sapiens] >spl	263	263	96%	4e-78	33%	NP_006835.2
2-hydroxyphytanoyl-CoA lyase [Homo sapiens]	229	229	96%	1e-65	28%	CAB60200.1
2-hydroxyacyl-CoA lyase 1 isoform a [Homo sapiens] >sp|Q9UJ83.2|HACL1_HUMAN RecName: Full=2-hydroxyacyl-CoA lyase 1; AltName: Full=2-hydroxypt	226	226	96%	9e-65	28%	NP_036392.2
HSPC279 [Homo sapiens]	226	226	96%	9e-65	28%	AAF28957.1
unnamed protein product [Homo sapiens]	205	205	96%	2e-57	27%	BAG63473.1
2-hydroxyacyl-CoA lyase 1 isoform c [Homo sapiens]	204	204	96%	5e-57	27%	NP_001271344.1
unnamed protein product [Homo sapiens]	204	204	95%	8e-57	30%	BAC86096.1
2-hydroxyacyl-CoA lyase 1 isoform b [Homo sapiens] >dbj|BAG63043.1| unnamed protein product [Homo sapiens]	193	193	96%	5e-53	26%	NP_001271342.1
ilvB (bacterial acetolactate synthase)-like isoform 1 variant [Homo sapiens]	163	163	53%	3e-43	31%	BAD92302.1
Acetolactate synthase [Homo sapiens]	159	159	64%	7e-42	33%	AAC18916.1
unnamed protein product [Homo sapiens]	154	154	62%	4e-40	29%	BAG61303.1
2-hydroxyacyl-CoA lyase 1 isoform d [Homo sapiens] >dbj|BAG63397.1| unnamed protein product [Homo sapiens]	153	206	78%	2e-39	29%	NP_001271345.1
ILVBL protein [Homo sapiens]	124	124	29%	3e-31	39%	AAI26914.1
IlvB (bacterial acetolactate synthase)-like [Homo sapiens]	107	107	22%	4e-25	44%	AAH00109.1
phytanoyl-CoA 2-hydroxylase 2, isoform CRA_b [Homo sapiens]	106	106	32%	6e-25	31%	EAW64252.1
phytanoyl-CoA 2-hydroxylase 2, isoform CRA_c [Homo sapiens] >dbj|BAG51836.1| unnamed protein product [Homo sapiens]	107	107	36%	1e-24	29%	EAW64253.1
HACL1 protein [Homo sapiens]	53.9	53.9	16%	1e-07	30%	AAH07440.1

Fig: 9 Analysis of the BLAST result in NCBI.

This page will show the identity and detail information under description column. Detail information and the most valuable file format of the gene can be achieved by marking the gene in the provided box then clicking on the download or GenPept, graphical view for Graphics. We can reduce the number of column by clicking on the setting icon on top of right corner.

All the BLAST result is listed below with their EC number and biological process:

	Accession No and Gene Name	Location in cell and cellular components	Can be used as target or not	Biological processes	Enzyme Commission Number
1.	Rv1820: ilvG	Not known	Yes	Amino acid biosynthesis	2.2.1.6
2.	Rv0951: sucC	Cytosol	No	Tricarboxylic acid cycle	6.2.1.5
3.	Rv0761c: adhb	Cytoplasm, plasma membrane	No	Oxidoreductase	1.1.1.1 1.1.1.284
4.	Rv0373c	Plasma membrane	Yes	Carbon monoxide dehydrogenase	1.2.99.2
5.	Rv0070c: glyA2	Cytoplasm	No	Serine hydroxymethyltransferase	2.1.2.1
6.	Rv1023: eno	Cytoplasm secreted	No	Tricarboxylic acid cycle	4.2.1.11
7	Rv1240: mdh	Cytosol, plasma membrane	No	Glycolysis	1.1.1.37
8.	Rv0363c: fba	Plasma membrane	Yes	Glycolysis, protein homotetramerization	4.1.2.13

9.	Rv2029c: pfkB	Not known	Yes	Carbohydrate metabolic process	2.7.1.11
10.	Rv0409: ackA	Cytoplasm	Yes	Organic acid metabolic process	2.7.2.1
11.	Rv0408:pta	Cytoplasm	No	Not known	2.3.1.8
12.	Rv3667:acs	Plasma membrane	No	Not known	6.2.1.1
13.	Rv0489:gpm1	Plasma membrane	No	phosphoglycerate mutase	5.4.2.11
14.	Rv0728c:serA2	Not known	Close to 35%, No	D-3-phosphoglycerate dehydrogenase	1.1.1.95
15.	Rv2611c:htrB	Plasma membrane	Yes	Acyltransferase	2.3.1.-
16.	Rv0113:gmhA	Cytoplasm	Yes	Carbohydrate metabolism	5.3.1.28
17.	Rv0114:gmhB	Cytoplasm	Yes	Carbohydrate metabolism, histadine biosynthesis	3.1.3.83 3.1.3.82
18.	Rv1315:murA	Cytoplasm	Yes	Cell cycle, cell division Peptidoglycan biosynthesis, regulation of cell shape	2.5.1.7
19.	Rv0482:murB	Cytoplasm	Yes	-do-	1.3.1.98
20.	Rv2152c:murC	Cytoplasm	Yes	-do-	6.3.2.8
21.	Rv2155c:murD	Cystol	Yes	-do-	6.3.2.9

22.	Rv2158c:murE	Plasma membrane	Yes	-do-	6.3.2.13
23.	Rv2981c:ddlA	Plasma membrane	Yes	-do-	6.3.2.4
24.	Rv2157c:murF	Cytoplasm	Yes	-do-	6.3.2.10
25.	Rv2136c	Plasma membrane	Yes	-do-, antibiotic resistance	3.6.1.27
26.	Rv2156c:murX	Plasma membrane	Yes	Cell cycle, cell division Peptidoglycan biosynthesis, regulation of cell shape	2.7.8.13
27.	Rv2153c:murG	Plasma membrane	Yes	-do-	2.4.1.227
28.	Rv3910	Plasma membrane	Yes	Conserved transmembrane protein	
29.	Rv0016c:pbpA	plasma membrane	Yes	penicillin-binding protein	2.4.1.129
30.	Rv2163c:pbpB	Extracellular	Yes	penicillin-binding membrane protein	

4.4 DISCUSSION:

The NCBI Gene ID of Rv 2981c is 888415, D-alanine--D-alanine ligase DdlA (D-alanylalanine synthetase) (D-ala-D-ala ligase). The catalytic activity of this protein is: ATP + 2 D-alanine = ADP + phosphate + D-alanyl-D-alanine. In this gene the magnesium binding site is present in the 318,330 and 332 of the amino acid sequence. This metal binding site is also important for this gene, if these site can be blocked then the biological function of the bacteria will stop and a important target can be achieved for treatment. The amino acid sequence of this gene is :

```
        10         20         30         40         50         60
MSANDRRDRR VRVAVVFGGR SNEHAISCVS AGSILRNLDS RRFDVIAVGI TPAGSWVLTD

        70         80         90        100        110        120
ANPDALTITN RELPQVKSGS GTELALPADP RRGGQLVSLP PGAGEVLESV DVVFPVLHGP

       130        140        150        160        170        180
YGEDGTIQGL LELAGVPYVG AGVLASAVGM DKEFTKKLLA ADGLPVGAYA VLRPPRSTLH

       190        200        210        220        230        240
RQECERLGLP VFVKPARGGS SIGVSRVSSW DQLPAAVARA RRHDPKVIVE AAISGRELEC

       250        260        270        280        290        300
GVLEMPDGTL EASTLGEIRV AGVRGREDSF YDFATKYLDD AAELDVPAKV DDQVAEAIRQ

       310        320        330        340        350        360
LAIRAFAAID CRGLARVDFF LTDDGPVINE INTMPGFTTI SMYPRMWAAS GVDYPTLLAT

       370
MIETTLARGV GL
```

This can be taken as target protein, because of the following points:

1. The 3D structure of the protein is known.
2. It has given the best BLASTP result.
3. This gene is responsible for the cell wall organization, regulation of cell shape. So, if gene is targeted the wall will disrupt and cell lysis will occur.
4. If this protein is checked, the metabolism followed after this gene will not occur and hence the metabolism product will not present in human.
5. As this gene is responsible for the cell wall organization, so the bacteria will not be able to go to dormant stage.

4.5 Structure of the gene Rv2981c:

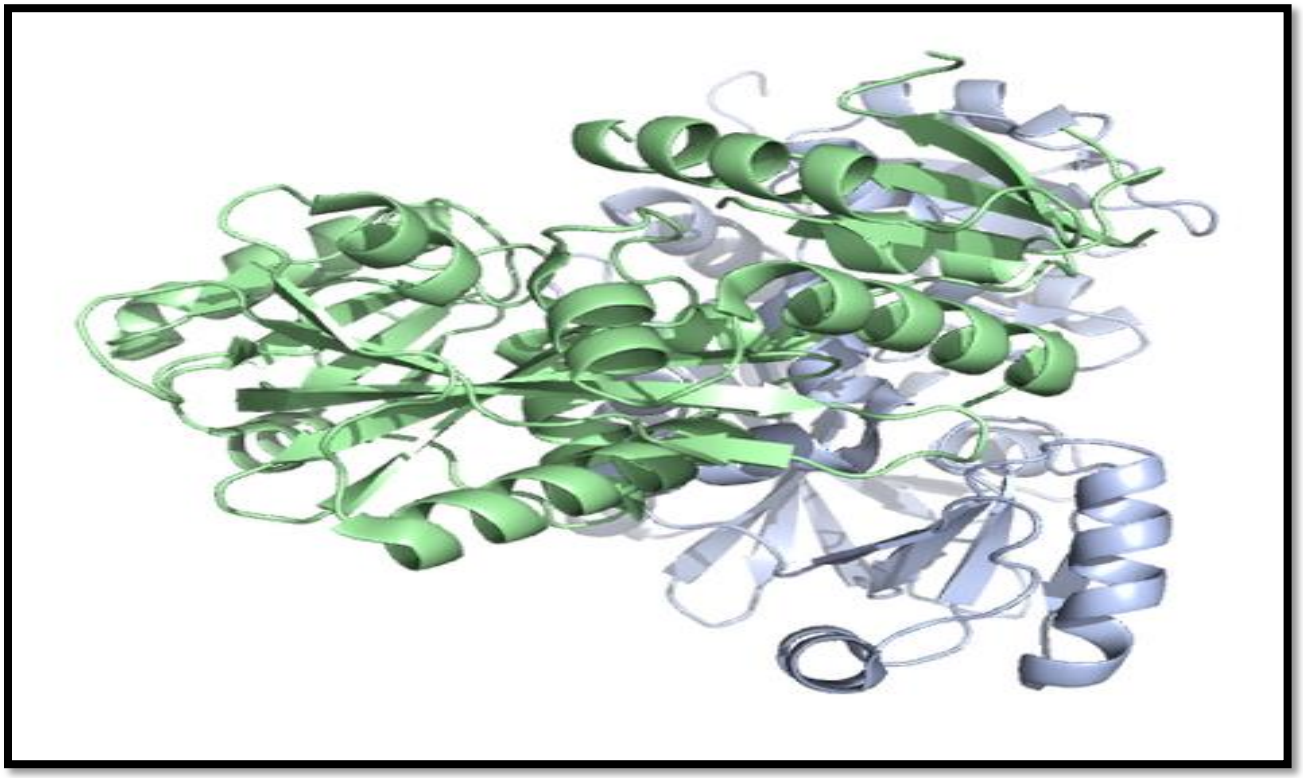


Fig: 10

5. CONCLUSION:

All the genes of the four important metabolism pathways which are not present in the human are taken as important pathways. The genes responsible for these pathways are listed and analyzed for their essentiality in the metabolism process of the bacteria through the DEG tool. All the shorted genes from the DEG are being shorted for the non-homologous for the human being through the NCBI BLAST. Here BLASTP is used, as BLASTP takes the amino acid sequence which is responsible for the protein present in the bacteria. And the introns will not present in our sequence. After performing all the above steps we have been able to find a target named as Rv2981c, which is present in the cytoplasm of the bacterial cell. If this gene function can be stopped then the bacteria can be killed and Multiple Drugs Resistant TB can be cured. As this gene is responsible for the cell shape, cell wall biogenesis/ degradation and peptidoglycan synthesis. This will disrupt the cell wall and will lead to the death of the bacteria.

The future scope of this project is to find a certain molecule which can be able to target this gene and the Multiple Drugs Resistant TB (MDR) can be cured.

6. REFERENCE:

1. Science Reporter, April, 2014.

2. Kaufmann SHE, "Envisioning future strategies for vaccination against tuberculosis", Nature Reviews immunology, vol.6, no.9,pp699-704,2006.

3. Nunn P, Williams B, Floyd K, Dye C, Elzinga G and Raviglione M. "Tuberculosis control in the era of HIV," Nature Reviews Immunology, vol 5, no. 10, pp. 819-826, 2005.

4. http://chealth.canoe.ca/index.asp?relation_id=3431.

5. Collins CH, Grange JM,Yates MD. Tuberculosis bacteriology. Organization and practice. 2nd edition Oxford: Butterworth Heinemann, 1997.

6. International Union of Tuberculosis and lung disease/world health organization. Statement: Tuberculosis preventive therapy in HIV infected individuals. Tuber lung disease 1994;75: 96-98.

7. Olson M, Hood L, Cantor C, Botstein D. A common language for pfysical mapping of human genome. Science. 1989;245(4925): 1434-1435.

8. Casey RM. (2005) "Blast Sequence Aid in Genomics and Proteomics". Business Intelligence Network.

9. Kanehisa M, Goto S. (2000), "KEGG: Kyoto Encyclopedia of Genes and Genomes". Nucleic Acids Res 28(1): 27-30.

10. Zhang RHY and Zhang CT. DEG: A database of essential genes. Nucleic acid research. 2004:32:D271-D272.

11. Uniprot C. (2010). "Ongoing future developments at the universal protein resource". Nucleic acid research 39 (Database issue): D214-D219.